# MALLIAVIN CALCULUS APPROACH TO STATISTICAL INFERENCE FOR LÉVY DRIVEN SDE'S

D. O. IVANENKO AND A. M. KULIK

Abstract. By means of the Malliavin calculus, integral representations for the likelihood function and for the derivative of the log-likelihood function are given for a model based on discrete time observations of the solution to equation $\mathrm{d}X_t = a_\theta(X_t)\mathrm{d}t + \mathrm{d}Z_t$ with a tempered $\alpha$-stable process $Z$. Using these representations, regularity of the statistical experiment and the Cramer-Rao inequality are proved.

## 1. Introduction

Consider stochastic equation of the form

$$\mathrm{d}X_t = a_\theta(X_t)\mathrm{d}t + \mathrm{d}Z_t, \tag{1}$$

where $Z$ is a one-dimensional Lévy process, $a : \Theta \times \mathbb{R} \to \mathbb{R}$ is a measurable function, $\Theta \subset \mathbb{R}$ is a parametric set. The main objective of our study is the statistical inference of the unknown parameter $\theta$ given the observations of the solution to this equation at the discrete time set $\{kh, k = 1, \dots, n\}$ with some given $h > 0$.

In the model where the solution to (1) is observed at a discrete time set, the likelihood function is highly implicit. In this paper, we develop an approach which makes it possible to control the properties of the likelihood and log-likelihood functions only in the terms of the objects involved in the model: the function $a_\theta(x)$, its derivatives, and the Lévy measure of the Lévy process $Z$. This approach is based on an appropriate version of the Malliavin calculus for a Poisson point measure.

The Malliavin calculus, developed first as a tool for proving existence and smoothness of distribution densities, appears to be very efficient in a study of sensitivities of expectations w.r.t. parameters. This field of applications, motivated by the analysis of volatilities in the models of financial mathematics, comes back to [6] and was studied intensively during the last years. An extension of this technique to statistical problems, similar to the one described above, looks quite natural. However, any publications are not available for us in this direction, with one important exception of the recent paper [5], where a Malliavin calculus-based sensitivity analysis was developed, and applications of this analysis were given to evaluation of the Cramer-Rao inequality and study of asymptotic properties of MLE for discretly observed diffusion processes.

We are mainly concentrated on the study of equation (1), where $Z$ is a Lévy process without a diffusion component. We develop a particular version of the Malliavin calculus for Poisson point measures from [2], [1], which is convenient for the purposes of the further sensitivity analysis. We give integral representations for the likelihood function and for the the derivative of the log-likelihood function

*Date*: Received: date / Accepted: date.

w.r.t. parameter. These representations then are used as a key ingredient in the proof of the regularity of a statistical experiment generated by discrete time set observations of the solution to (1), and consequent evaluation of the Cramer-Rao inequality.

In this paper, we confine ourselves to the case of a Lévy process $Z$ generated by a tempered $\alpha$-stable Poisson measure ($T\alpha S$ in the sequel, see [12]). Such processes arise naturally in models of turbulence [10], economical models of stochastic volatility [3], etc. Generalization of the results of this paper for equations (1) with other types of noises, multidimensional equations, and so on, and application of these results to study the asymptotic properties of Maximal Likelihood Estimator is a subject of further research.

## 2. Notation, assumptions, and main results

2.1. **Notation and assumptions.** Let $Z$ be a Lévy process without a diffusion component; that is,

$$Z_t = ct + \int_0^t \int_{|u|>1} u\nu(\mathrm{d}s, \mathrm{d}u) + \int_0^t \int_{|u|\leq 1} u\tilde{\nu}(\mathrm{d}s, \mathrm{d}u),$$

where $\nu$ is a Poisson point measure with the intensity measure $\mathrm{d}s\mu(\mathrm{d}u)$, and $\tilde{\nu}(\mathrm{d}s, \mathrm{d}u) = \nu(\mathrm{d}s, \mathrm{d}u) - \mathrm{d}s\mu(\mathrm{d}u)$ is respective compensated Poisson measure. In the sequel, we assume $Z$ to be a *tempered $\alpha$-stable* process, which means that

$$\mu(\mathrm{d}u) = |u|^{-\alpha-1}\Big(C_-\mathbf{1}_{u<0} + C_+\mathbf{1}_{u>0}\Big)r(u)\mathrm{d}u \tag{2}$$

with some $C_\pm \geq 0$ and some function $r$ such that $r(0) > 0$. We assume that $r \in C^2(\mathbb{R})$, $r'(0) = 0$, and $r, r', r''$ tend to zero as $|x| \to \infty$ faster than any power of $1/|x|$.

For a given $\theta$, assuming that the drift term $a_\theta$ satisfies the standard local Lipschitz and linear growth conditions, Eq. (1) uniquely defines a Markov process $X$. We denote by $\mathsf{P}_x^\theta$ the distribution of this process in $\mathbb{D}([0,\infty))$ with $X_0 = x$, and by $\mathsf{E}_x^\theta$ the expectation w.r.t. this distribution. Respective finite-dimensional distribution for given time moments $t_1 < \cdots < t_n$ is denoted by $\mathsf{P}^\theta_{x,\{t_k\}_{k=1}^n}$.

In the sequel we will show that, under appropriate conditions, Markov process $X$ admits a transition probability density $p_t^\theta(x,y)$ w.r.t. Lebesgue measure, which is continuous w.r.t. $(t,x,y) \in (0,\infty)\times\mathbb{R}\times\mathbb{R}$. Then (see [4]), for every $t > 0, x, y \in \mathbb{R}$ such that

$$p_t^\theta(x,y) > 0, \tag{3}$$

there exists a weak limit in $\mathbb{D}([0,t])$

$$\mathsf{P}^{t,\theta}_{x,y} = \lim_{\varepsilon\to 0} \mathsf{P}_x^\theta\Big(\,\cdot\,\Big|\,|X_t - y| \leq \varepsilon\Big),$$

which can be interpreted naturally as a *bridge* of the process $X$ started at $x$ and conditioned to arrive to $y$ at time $t$. We denote by $\mathsf{E}^{t,\theta}_{x,y}$ the expectation w.r.t. $\mathsf{P}^{t,\theta}_{x,y}$.

In what follows, $C$ denotes a constant which is not specified explicitly and may vary from place to place. By $C^{k,m}(\mathbb{R}\times\Theta), k, m \geq 0$ we denote the class of functions $f : \mathbb{R}\times\Theta \to \mathbb{R}$ which has continuous derivatives

$$\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial \theta^j}f, \quad i \leq k, \quad j \leq m.$$

2.2. **Main results: formulation.** Here we formulate two main theorems of this paper. The first one concerns the local properties of the transition probabilities of the Markov process $X$. The functionals $\Xi_t,\ \Xi_t^1$, involved in its formulation, will be introduced explicitly in the proof below; see formulae (32) and (35).

**Theorem 1.** *I. Let $a \in C^{2,0}(\mathbb{R}\times\Theta)$ with bounded derivatives $\partial_x a_\theta, \partial^2_{xx} a_\theta$.*

*Then the Markov process $X$ given by (1) has a transition probability density $p_t^\theta$ w.r.t. the Lebesgue measure, which has an integral representation*

$$p_t^\theta(x,y) = \mathsf{E}_x^\theta \left[\Xi_t \mathbb{1}_{X_t>y}\right], \quad t>0, \quad x,y\in\mathbb{R}. \tag{4}$$

*The function $p_t^\theta(x,y)$ is continuous w.r.t. $(t,x,y,\theta)\in(0,\infty)\times\mathbb{R}\times\mathbb{R}\times\Theta$.*

*II. Let $a\in C^{3,1}(\mathbb{R}\times\Theta)$ have bounded derivatives $\partial_x a$, $\partial^2_{xx}a$, $\partial^2_{x\theta}a$, $\partial^3_{xx\theta}a$, $\partial^4_{xxx\theta}a$ and*

$$|a_\theta(x)| + |\partial_\theta a_\theta(x)| \le C(1+|x|), \quad \theta\in\Theta, \quad x\in\mathbb{R}. \tag{5}$$

*Then the transition probability density has a derivative $\partial_\theta p_t^\theta(x,y)$, which is continuous w.r.t. $(t,x,y,\theta)\in(0,\infty)\times\mathbb{R}\times\mathbb{R}\times\Theta$.*

*III. Under the conditions of statement II, one has*

$$\partial_\theta p_t^\theta(x,y) = g_t^\theta(x,y) p_t^\theta(x,y) \tag{6}$$

*with*

$$g_t^\theta(x,y) = \begin{cases} \partial_\theta \log p_t^\theta(x,y) = \mathsf{E}^{t,\theta}_{x,y}\Xi_t^1, & p_t^\theta(x,y)>0,\\ 0, & \textit{otherwise}.\end{cases} \tag{7}$$

**Remark 1.** *By statements II and III, the logarithm of the transition probability density has a continuous derivative w.r.t. $\theta$ on the open subset of $(0,\infty)\times\mathbb{R}\times\mathbb{R}\times\Theta$ defined by inequality (3) and, on this subset, admits the integral representation*

$$\partial_\theta \log p_t^\theta(x,y) = \mathsf{E}^{t,\theta}_{x,y}\Xi_t^1. \tag{8}$$

The second theorem concerns the basic properties of the statistical experiment

$$\Big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathsf{P}^\theta_{x,\{t_k\}_{k=1}^n}, \theta\in\Theta\Big), \tag{9}$$

generated by observations of the Markov process $X$ with fixed $X_0=x$ at time moments $t_1<\dots<t_n$; we refer to [8] for the notation and terminology. Recall that a statistical experiment $(\mathcal{X},\mathcal{U},P^\theta,\theta\in\Theta)$ is called *regular*, if $\mathrm{d}P^\theta = p^\theta \mathrm{d}\lambda$ for some $\sigma$-finite measure $\lambda$, and

(a) the function $\theta\mapsto p^\theta(\mathbf{x})$ is continuous for $\lambda$-a.a. $\mathbf{x}\in\mathcal{X}$;

(b) the function $\theta\mapsto\sqrt{p^\theta}\in L_2(\mathcal{X},\lambda)$ is continuously differentiable.

For a regular statistical experiment with $\Theta\subset\mathbb{R}^1$, respective *Fisher information* is defined as

$$I(\theta) = 4\int_{\mathcal{X}}\Big(\partial_\theta\sqrt{p^\theta}\Big)^2 \mathrm{d}\lambda,$$

with the derivative understood in the $L_2(\mathcal{X},\lambda)$ sense.

**Theorem 2.** *Let conditions of statement II of Theorem hold true and $x\in\mathbb{R}, n\in\mathbb{N}$, $0<t_1<\dots<t_n$ be fixed.*

*Then the statistical experiment (9) is regular. Respective Fisher information equals*

$$I(\theta) = \sum_{k=1}^n E_x^\theta\Big(g_h^\theta(X_{t_{k-1}}, X_{t_k})\Big)^2,$$

*where* $t_0 := 0$.

**Remark 2.** *For the statistical experiment (9)* $\mathcal{X} = \mathbb{R}^n$, *and the natural choice of* $\lambda$ *is the Lebesgue measure. Then by Theorem 1*

$$p^\theta(\mathbf{x}) = \prod_{k=1}^n p^\theta_{t_k - t_{k-1}}(x_{k-1}, x_k), \quad \mathbf{x} = (x_1, \dots, x_n), \tag{10}$$

*where* $x_0 := x$, *and there exists a point-wise derivative*

$$\partial_\theta p^\theta(\mathbf{x}) = p^\theta(\mathbf{x}) g^\theta(\mathbf{x}), \quad g^\theta(\mathbf{x}) := \sum_{k=1}^n g^\theta_{t_k - t_{k-1}}(x_{k-1}, x_k), \quad \mathbf{x} = (x_1, \dots, x_n). \tag{11}$$

*In particular, one can interpret (4) and (8) as integral representations for the likelihood function and the derivative of the likelihood function in a one-point observation model.*

Combining Theorem 2 and Theorem I.7.3 in [8], we obtain the following version of the Cramer-Rao inequality.

**Corollary 1.** *Let conditions of statement II of Theorem hold true and* $x \in \mathbb{R}, n \in \mathbb{N}$, $0 < t_1 < \dots < t_n$ *be fixed. Assume that*

$$I(\theta) > 0$$

*for some* $\theta$ *and* $T : \mathbb{R}^n \to \mathbb{R}$ *is a Borel measurable function such that the function*

$$\theta \mapsto \mathsf{E}^\theta_x T^2(X_{t_1}, \dots, X_{t_n})$$

*is locally bounded.*

*Then the bias*

$$d(\theta) = \mathsf{E}^\theta_x T(X_{t_1}, \dots, X_{t_n}) - \theta$$

*is differentiable, and*

$$\mathsf{E}^\theta_x \Big( T(X_{t_1}, \dots, X_{t_n}) - \theta \Big)^2 \geq \frac{(1 + \partial_\theta d(\theta))^2}{I(\theta)} + d^2(\theta).$$

2.3. **Main results: discussion.** Let us emphasize the point at which our approach differ substantially from the one developed in the closely related paper [5]. Among others, the approach of [5] requires the following structural assumption:

$$\text{the support of the density } p^\theta \text{ does not depend on } \theta. \tag{12}$$

Under the conditions of Theorem 1, statement II, assumption (12) would simplify the subsequent statistical analysis drastically, because one could restrict then the state space $\mathcal{X}$ to the respective support and obtain the model with $C^1$ log-likelihood function. However, an attempt to *prove* that in the model defined by Eq. (1) the density (10) satisfies (12), meets serious difficulties. Moreover, one can see from the example below that, without further restrictions on the model, this assumption fails at all.

**Example 1.** *Let* $\alpha < 1$ *and* $Z$ *has positive jumps, only; that is,* $C_- = 0$ *and* $C_+ > 0$. *Then by the support theorem from* [14], *the topological support of* $P^\theta_t(x, \mathrm{d}y)$ *equals* $[y^\theta_t(x), \infty)$, *where* $y^\theta_t(x)$ *is the value at time moment* $s = t$ *of the solution to the Cauchy problem*

$$y'(s) = a_\theta(y(s)), \quad y(0) = x.$$

*Generically, $y^\theta_t(x)$ depends on $\theta$; for instance, for $a_\theta(x) = \theta x$ one has $y^\theta_t(x) = xe^{t\theta}$. Because the topological support of $P^\theta_t(x, \mathrm{d}y)$ is the closure of the support of the transition probability density $\mathrm{p}^\theta_t(x, y)$, this indicates that in that case (12) fails.*

This observation mainly motivates the particular form of our approach: we would like to exclude the assumption (12) completely, and therefore we do not rely on the path-wise regularity of the log-likelihood function. Instead of that, we show that our model can be embedded naturally into the general framework of regular statistical experiments; e.g. [8].

## 3. Malliavin calculus for $T\alpha S$ random measures

Typically, a Malliavin calculus-based sensitivity analysis requires a pair of a derivation operator $D$ and an adjoint operator $\delta = D^*$ to be defined on the probability space under the consideration. Below we outline the construction of such operators for the particular case of a probability space with a $T\alpha S$ random measure. Such a construction, based on perturbations of "jump amplitudes", is well known in the field and goes back to [2], [1]. To keep the exposition self-sufficient and explicit, we explain the main components of this construction important for further statistical applications.

3.1. **Perturbations of $T\alpha S$ random measures and associated differential operators.** Denote by $\mathcal{O}$ the space of *locally finite configurations* in $\mathbb{R}^+ \times (\mathbb{R} \setminus \{0\})$, endowed by the *vague topology*; for details, see e.g. [9]. This is the natural state space when the random point measure $\nu$ is considered as a random element; denote by $\mathsf{P}_\nu$ the distribution of $\nu$ in $(\mathcal{O}, \mathcal{B}(\mathcal{O}))$. In what follows, we identify the basic probability space $(\Omega, \mathcal{F}, \mathsf{P})$ with $(\mathcal{O}, \mathcal{B}(\mathcal{O}), \mathsf{P}_\nu)$, and assume $\nu(\omega) = \omega$.

Let $\varrho : \mathbb{R} \to \mathbb{R}^+$ be a Lipschitz continuous function. Denote by $Q_c(x), c \in \mathbb{R}$ the value at time moment $s = c$ of the solution to Cauchy problem

$$q'(s) = \varrho(q(s)), \quad q(0) = x.$$

Then $\{Q_c, c \in \mathbb{R}\}$ is a group of transformations of $\mathbb{R}$, and $\partial_c Q_c(x)|_{c=0} = \varrho(x)$.

For given $T > 0$ and $\varrho$, define the group $\{\mathcal{Q}_c, c \in \mathbb{R}\}$ of transformations of the configuration space $\mathcal{O}$ in the following way. Every $\omega \in \mathcal{O}$ is a locally finite collection of points $(\tau, u)$, where $\tau \in \mathbb{R}^+$ is the "jump time", and $u \in \mathbb{R} \setminus \{0\}$ is the "jump amplitude". Transformation $\mathcal{Q}_c$ maps a configuration $\omega$ into the collection of points of the form

$$\begin{cases} (\tau, Q_c(u)), & (\tau, u) \in \omega \text{ is such that } \tau \le T; \\ (\tau, u), & (\tau, u) \in \omega \text{ is such that } \tau > T. \end{cases}$$

One can see easily that $\mathcal{Q}_c$ transforms $\nu$ into the Poisson point measure $\nu_c$ s.t.

$$\nu_c(A \times B) = \nu\left((A \cap [0, T]) \times Q_c^{-1}(B)\right) + \nu\left(\left(A \cap \overline{[0, t]}\right) \times B\right)$$

for every $A \in \mathcal{B}(\mathbb{R}_+)$ and $B \in \mathcal{B}(\mathbb{R})$. An explicit calculation shows that the intensity measure for $\nu_c$ has the form

$$\mathbf{1}_{s \le T} \mathrm{d}s\, [\mu \circ Q_c^{-1}](\mathrm{d}u) + \mathbf{1}_{s > T} \mathrm{d}s\, \mu(\mathrm{d}u). \tag{13}$$

Fix $u_0 > 0$ such that $\inf_{|u| \le 2u_0} r(u) > 0$, where $r(u)$ comes from (2). In what follows, the function $\varrho$ involved in the definition of $Q_c, c \in \mathbb{R}$ is assumed to satisfy

$$\varrho(u) = \begin{cases} |u|^{\beta+1}, & |u| \le u_0; \\ 0, & |u| \ge 2u_0 \end{cases}$$

with some $\beta \geq 0$. Then the intensity measure (13) has the density w.r.t. $\mathrm{d}s\mu(\mathrm{d}u)$ equal to

$$m_{c,T}(s,u) = \mathbb{1}_{s\leq T} m_c(u) + \mathbb{1}_{s>T}, \tag{14}$$

where

$$m_c(u) = \frac{\mathrm{d}[\mu \circ Q_c]^{-1}}{\mathrm{d}\mu}(u) = \frac{1}{R_c(Q_c^{-1}(u))} \frac{|Q_c^{-1}(u)|^{-\alpha-1} r(Q_c^{-1}(u))}{|u|^{-\alpha-1} r(u)}$$

with

$$R_c(x) := \partial_x Q_c(x) = \exp\left(\int_0^c \varrho'(Q_s(x))\,\mathrm{d}s\right).$$

Straightforward calculation shows that $\log m_c$ is a continuous function, which vanishes when $|u| \geq u_0$, and satisfies

$$\log m_c(u) \sim C|u|^\beta, \quad u \to 0.$$

Therefore, when $\beta$ satisfies

$$\beta > \alpha/2, \tag{15}$$

one has

$$\int\limits_{|\log m_{c,T}|\geq \log 2} |1 - m_{c,T}(s,u)|\,\mathrm{d}s\mu(\mathrm{d}u) + \int\limits_{\mathbb{R}^+\times\mathbb{R}} \frac{\log^2 m_{c,T}(s,u)}{1+\log^2 m_{c,T}(s,u)} \Pi(\mathrm{d}u) < \infty.$$

Applying Skorokhod's criterion for absolute continuity of the laws of Poisson point measures [15], we arrive at following.

**Proposition 1.** *Let $\beta$ satisfy (15).*

*Then the distribution $\mathsf{P}_{\nu_c}$ of $\nu_c$ in $(\mathcal{O}, \mathcal{B}(\mathcal{O}))$ is absolutely continuous w.r.t. $\mathsf{P}_\nu$, and*

$$\kappa_c := \frac{\mathrm{d}\mathsf{P}_{\nu_c}}{\mathrm{d}\mathsf{P}_\nu}(\nu) = \exp\left\{\int_0^T \int_{\mathbb{R}} \log m_c(u)\tilde{\nu}(\mathrm{d}s, \mathrm{d}u) + T\int_{\mathbb{R}} (1 - m_c(u) + \log m_c(u))\,\mu(\mathrm{d}u)\right\}. \tag{16}$$

Consequently, the map $\mathcal{Q}_c : \Omega \to \Omega$ generates the map of $L_0(\Omega, \mathcal{F}, \mathsf{P})$ into itself in the following way (we keep the same symbol $\mathcal{Q}_c$ for this map):

$$\mathcal{Q}_c F(\nu) = F(\mathcal{Q}_c \nu), \quad F \in L_0(\Omega, \mathcal{F}, \mathsf{P}).$$

Straightforward computation shows that for every $u \neq 0$

$$\partial_c m_c(u)|_{c=0} = -\frac{(|u|^{-\alpha-1} r(u)\varrho(u))'}{|u|^{-\alpha-1} r(u)} =: \chi(u),$$

and under the assumption (15)

$$\int_{\mathbb{R}} \left(\frac{m_c(u) - 1}{c} - \chi(u)\right)^2 \mu(\mathrm{d}u) \to 0, \quad c \to 0.$$

Because $m_c(u) = 1, |u| \geq u_0$, the latter relation and (16) yield

$$\frac{\kappa_c(u) - 1}{c} \to \int_0^T \int_{\mathbb{R}} \chi(u)\tilde{\nu}(\mathrm{d}s, \mathrm{d}u) \text{ in every } L_p(\Omega, \mathcal{F}, \mathsf{P}),\ p \geq 1. \tag{17}$$

The proof of (17) is straightforward but cumbersome, and therefore is omitted.

**Definition 1.** *A functional $F \in L_2(\Omega, \mathcal{F}, \mathsf{P})$ is called* stochastically differentiable, *if there exists an $L_2(\Omega, \mathcal{F}, \mathsf{P})$-limit*

$$\hat{\mathrm{D}}F = \lim_{c \to 0} \frac{1}{c}\Big(\mathcal{Q}_c F - F\Big). \tag{18}$$

*The closure* D *of the operator* $\hat{\mathrm{D}}$ *defined by (18) is called the* stochastic derivative. *The adjoint operator $\delta = \mathrm{D}^*$ is called the* divergence operator *or the* extended stochastic integral.

**Remark 3.** *By (19) and (24) below,* dom(D) *is dense in $L_2(\Omega, \mathcal{F}, \mathsf{P})$, hence $\delta$ is well defined. In addition, by statement 3 of Proposition 2 below* dom($\delta$) *is dense in $L_2(\Omega, \mathcal{F}, \mathsf{P})$, hence $\hat{\mathrm{D}}$ is closable. The operator $\delta$ itself is closed as an adjoint one; e.g. Theorem VIII.1 in* [13].

The following proposition collects the main properties of the operators $\mathrm{D}, \delta$. In what follows we assume (15) to hold true.

**Proposition 2.** *1. Let $\varphi \in C^1(\mathbb{R}^d, \mathbb{R})$ have bounded derivatives and $F_k \in \mathrm{dom}(\mathrm{D})$, $k = \overline{1, d}$.*

*Then $\varphi(F_1, \dots, F_d) \in \mathrm{dom}(\mathrm{D})$ and*

$$\mathrm{D}\left[\varphi(F_1, \dots, F_d)\right] = \sum_{k=1}^d [\partial_{x_k}\varphi](F_1, \dots, F_d)\mathrm{D}F_k. \tag{19}$$

*2. The constant function* 1 *belongs to* dom($\delta$) *and*

$$\delta(1) = \int_0^T \int_{\mathbb{R}} \chi(u)\tilde{\nu}(\mathrm{d}s, \mathrm{d}u). \tag{20}$$

*3. Let $G \in \mathrm{dom}(\mathrm{D})$ and*

$$\mathsf{E}\left(\delta(1)G\right)^2 < \infty. \tag{21}$$

*Then $G \in \mathrm{dom}(\delta)$ and $\delta(G) = \delta(1)G - \mathrm{D}G$.*

*Proof.* 1. It is sufficient to consider $F_k, k = \overline{1, d}$ which satisfy (18). Then the fraction

$$\frac{1}{c}(\mathcal{Q}_c\varphi(F_1, \dots, F_d) - \varphi(F_1, \dots, F_d)) \tag{22}$$

converges in probability to the right hand side of (19). Its square is dominated by

$$\sup_x \|\nabla\varphi(x)\|^2 \sum_{k=1}^d \left(\frac{\mathcal{Q}_c F_k - F_k}{c}\right)^2,$$

and hence is uniformly integrable. Consequently, (22) converges in $L_2(\Omega, \mathcal{F}, \mathsf{P})$.

2. For any $F$ satisfying (18) we have by (17) with $p = 2$

$$\mathsf{E}F\left(\int_0^T \int_{\mathbb{R}} \chi(u)\tilde{\nu}(\mathrm{d}s, \mathrm{d}u)\right) = \lim_{c \to 0} \frac{1}{c}\mathsf{E}F(\kappa_c - 1) = \lim_{c \to 0} \frac{1}{c}\mathsf{E}(\mathcal{Q}_c F - F) = E\mathrm{D}F,$$

which gives by the definition of $\delta = \mathrm{D}^*$ that $\delta(1) = \int_0^T \int_{\mathbb{R}} \chi(u)\tilde{\nu}(\mathrm{d}s, \mathrm{d}u)$.

3. For *bounded* $F, G \in \mathrm{dom}(\mathrm{D})$ one has by (19) $FG \in \mathrm{dom}(\mathrm{D})$ and $\mathrm{D}(FG) = F\mathrm{D}G + G\mathrm{D}F$. Then by statement 2 we have

$$\mathsf{E}G\mathrm{D}F = \mathsf{E}F\Big(\delta(1)G - \mathrm{D}G\Big). \tag{23}$$

For arbitrary $F \in \mathrm{dom}(\mathrm{D})$, using (19), one can choose a sequence of bounded $F_n \in \mathrm{dom}(\mathrm{D})$ such that $F_n \to F$ and $\mathrm{D}F_n \to \mathrm{D}F$ in $L_2(\Omega, \mathcal{F}, \mathsf{P})$. This proves (23) for arbitrary $F \in \mathrm{dom}(\mathrm{D})$, and yields the required statement under the additional assumption that $G$ is bounded. Approximating $G$ by bounded $G_n \in \mathrm{dom}(\mathrm{D})$ and using that $\delta$ is a closed operator completes the proof. $\square$

3.2. **Differential properties of the solution to** (1)**.** Denote $Z_t^c = \mathcal{Q}_c Z_t$. It can be seen straightforwardly that

$$\frac{1}{c}(Z_t^c - Z_t) \to \mathrm{D}Z_t := \int_0^t \int_{\mathbb{R}} \varrho(u)\nu(\mathrm{d}s, \mathrm{d}u) \text{ in every } L_p(\Omega, \mathcal{F}, \mathsf{P}),\ p \geq 1 \tag{24}$$

uniformly by $t \in [0, T]$ for every $T$. Then one can consider $X_t^c = \mathcal{Q}_c X_t$ as the solution to the following perturbed SDE:

$$\mathrm{d}X_t^c = a_\theta(X_t^c)\mathrm{d}t + \mathrm{d}Z_t^c. \tag{25}$$

Applying Theorem 1.2.7.4 in [7], under conditions of statement I, Theorem 1 we get that for any fixed $\theta \in \Theta$ and initial value $x \in \mathbb{R}$

$$\frac{1}{c}(X_t^c - X_t) \to \mathrm{D}X_t \text{ in every } L_p(\Omega, \mathcal{F}, \mathsf{P}),\ p \geq 1 \tag{26}$$

uniformly by $t \in [0, T]$ for every $T$. The process $Y_t := \mathrm{D}X_t$ satisfies the linear SDE

$$\mathrm{d}Y_t = \partial_x a_\theta(X_t)\mathrm{d}s + \mathrm{d}\mathrm{D}Z_t, \quad Y_0 = 0,$$

and hence can be written explicitly:

$$\mathrm{D}X_t = \int_0^t \int_{\mathbb{R}} r(s,t)\varrho(u)\nu(\mathrm{d}s, \mathrm{d}u), \quad r(s,t) := \exp\left\{\int_s^t \partial_x a_\theta(X_\tau)\mathrm{d}\tau\right\}. \tag{27}$$

The same argument leads to the following formulae (we omit the detailed exposition):

$$\mathrm{D}r(s,t) = r(s,t)\int_s^t \partial^2_{xx} a_\theta(X_\tau)\mathrm{D}X_\tau \mathrm{d}\tau, \tag{28}$$

$$\begin{aligned} \mathrm{D}^2 X_t := \mathrm{D}(\mathrm{D}X_t) &= \int_0^t \mathrm{D}r_\theta(s,t)\mathrm{d}\mathrm{D}Z_s + \int_0^t r_\theta(s,t)\mathrm{d}\mathrm{D}^2 Z_s \\ &= \int_0^t \int_{\mathbb{R}} \Big(\mathrm{D}r(s,t)\varrho(u) + r(s,t)\varrho'(u)\varrho(u)\Big)\nu(\mathrm{d}s, \mathrm{d}u). \end{aligned} \tag{29}$$

Similarly, under conditions of statement II, Theorem 1 the solution to Eq. (1) is $L_p$-differentiable w.r.t. $\theta$ for every $p \geq 1$. Respective derivative equals

$$\partial_\theta X_t = \int_0^t r(s,t)\partial_\theta a_\theta(X_s)\mathrm{d}s \tag{30}$$

and is stochastically differentiable with

$$\mathrm{D}(\partial_\theta X_t^\theta) = \int_0^t \partial_\theta a_\theta(X_s^\theta)\mathrm{D}r(s,t)\mathrm{d}s + \int_0^t r(s,t)\partial^2_{x\theta} a_\theta(X_s)(\mathrm{D}X_s)\mathrm{d}s. \tag{31}$$

## 4. Proof of Theorem 1

First, we prove that $(\mathrm{D}X_t)^{-1} \in \mathrm{dom}(\delta)$ and

$$\Xi_t := \delta\left(\frac{1}{\mathrm{D}X_t}\right) = \frac{\delta(1)}{\mathrm{D}X_t} + \frac{\mathrm{D}^2 X_t}{(\mathrm{D}X_t)^2}. \tag{32}$$

This proof can be made in quite a standard way (see [11] for similar results in Gaussian setting), using Proposition 2 and the following moment bound for $(\mathrm{D}X_t)^{-1}$.

**Lemma 1.** *There exist $c, C > 0$ such that for every $x \in \mathbb{R}, \theta \in \Theta, t \le T$:*

$$\mathsf{E}_x^\theta \exp\left(ct(\mathrm{D}X_t)^{-\alpha/(\beta+1)}\right) \le C. \tag{33}$$

**Remark 4.** *Recall that $\varrho \ge 0$, and therefore $\mathrm{D}X_t \ge 0$.*

*Proof.* Because $\partial_x a_\theta$ is bounded, there exist positive $C_1, C_2$ such that

$$C_1 \le r(s,t) \le C_2, \quad s \le t \le T.$$

Then it is sufficient to prove the bound similar to (33) with $\mathrm{D}X_t$ replaced by

$$\eta_t := \int_0^t \int_{\mathbb{R}} \varrho(u)\nu(\mathrm{d}s, \mathrm{d}u) \ge \int_0^t \int_{|u|\le u_0} |u|^{\beta+1}\nu(\mathrm{d}s, \mathrm{d}u).$$

Denote $r_0 = \inf_{|u|\le u_0} r(u)$, which is positive by assumption, and put

$$u_k = k^{-1/\alpha}u_0, \quad \Delta_k = [u_{k-1}, u_k], \quad k \ge 1.$$

Then $\lambda_k := \mu(\Delta_k) \ge (r_0/\alpha)u_0^{-\alpha}$, and therefore

$$\eta_t \ge \sum_{k=1}^\infty k^{-(\beta+1)/\alpha}\eta_{t,k}, \quad \eta_{t,k} := \nu([0,t]\times\Delta_k), \quad k \ge 1.$$

For $\varepsilon > 0$, put $K_\varepsilon = \max\{k : k^{-(\beta+1)/\alpha} > \varepsilon\}$. Because $\{\eta_{k,t}\}$ are independent Poisson distributed r.v.'s with parameters $t\lambda_k$, we get finally

$$\begin{aligned}\mathsf{P}(\eta_t < \varepsilon) &\le \prod_{k=1}^{K_\varepsilon} \mathsf{P}(\eta_{t,k} = 0) \le e^{-t(r_0/\alpha)u_0^{-\alpha}k_\varepsilon} \\ &\le \exp\left\{-t(r_0/\alpha)u_0^{-\alpha}(\varepsilon^{-\alpha/(\beta+1)} - 1)\right\}, \quad \varepsilon > 0,\end{aligned}$$

which proves (33) for $\eta_t$. □

Now we can finalize the proof of *statement I*; the argument here is quite analogous to the one from [11], Chapter 3.1, hence we omit details. By Proposition 2, the definition of $\delta = \mathrm{D}^*$, and (32), for every $\varphi \in C_b^1(\mathbb{R})$ we have

$$\mathsf{E}_x^\theta\varphi'(X_t) = \mathsf{E}_x^\theta \mathrm{D}(\varphi(X_t))\left(\frac{1}{\mathrm{D}X_t}\right) = \mathsf{E}_x^\theta\varphi(X_t)\delta\left(\frac{1}{\mathrm{D}X_t}\right) = \mathsf{E}_x^\theta\varphi(X_t)\Xi_t.$$

Approximating $\varphi_y := \mathbf{1}_{[0,\infty)}(\cdot - y)$ by a sequence of $\varphi_n \in C_b^1(\mathbb{R})$, we get the representation (4). Clearly, both $X_t$ and $\Xi_t$ are continuous in probability w.r.t. parameters $x, t, \theta$. Because, by representation (4),

$$\mathsf{P}_x^\theta(X_t = y) = 0, \quad x, y \in \mathbb{R}, \quad t > 0, \quad \theta \in \Theta,$$

this proves that $p_t^\theta(x,y)$ is continuous w.r.t. $(t, x, y, \theta)$.

To prove *statement II*, we make one more integration by parts in the right hand side of (4). To do that, we use Proposition 2 and Lemma 1 and show that $\Xi_t/(\mathrm{D}X_t)$ belongs to $\mathrm{dom}(\delta)$ with

$$\delta\left(\frac{\Xi_t}{\mathrm{D}X_t}\right) = \frac{(\delta(1))^2 - \mathrm{D}\delta(1)}{(\mathrm{D}X_t)^2} + \frac{3\delta(1)\mathrm{D}^2X_t - \mathrm{D}^3X_t}{(\mathrm{D}X_t)^3} + \frac{3(\mathrm{D}^2X_t)^2}{(\mathrm{D}X_t)^4}; \tag{34}$$

here $\mathrm{D}\delta(1) = \int_0^T\int_{\mathbb{R}} \chi'(u)\varrho(u)\nu(\mathrm{d}s, \mathrm{d}u)$, and the formula for $\mathrm{D}^3X_t$ (cumbersome, but quite analogous to (29)) is omitted because it is not used explicitly in the sequel. Then (4) gives

$$p_t^\theta(x,y) = \mathsf{E}_x^\theta \psi_y(X_t)\delta\left(\frac{\Xi_t}{\mathrm{D}X_t}\right),$$

where $\psi_y = (\cdot - y)\vee 0$ is an absolutely continuous function with the derivative equal to $\varphi_y$. Note that both $X_t$ and $\delta\left(\Xi_t/(\mathrm{D}X_t)\right)$ are $L_p$-differentiable w.r.t. parameter $\theta$ for every $p \geq 1$. Hence there exists continuous derivative

$$\partial_\theta p_t^\theta(x,y) = \mathsf{E}_x^\theta\left[\varphi_y(X_t)\partial_\theta X_t \delta\left(\frac{\Xi_t}{\mathrm{D}X_t}\right) + \psi_y(X_t)\partial_\theta\delta\left(\frac{\Xi_t}{\mathrm{D}X_t}\right)\right].$$

To prove *statement III*, we again use Proposition 2 and Lemma 1 in order to show that $\partial_\theta X_t/(\mathrm{D}X_t)$ belongs to $\mathrm{dom}(\delta)$ and

$$\Xi_t^1 := \delta\left(\frac{\partial_\theta X_t}{\mathrm{D}X_t}\right) = \frac{(\partial_\theta X_t)\delta(1)}{\mathrm{D}X_t} + \frac{(\partial_\theta X_t)\mathrm{D}^2X_t}{(\mathrm{D}X_t)^2} - \frac{\mathrm{D}(\partial_\theta X_t)}{\mathrm{D}X_t}. \tag{35}$$

Then for any test function $f \in C^1(\mathbb{R})$ with a bounded derivative we have

$$\begin{aligned}\partial_\theta \mathsf{E}_x^\theta f(X_t^\theta) = \mathsf{E}_x^\theta f'(X_t^\theta)(\partial_\theta X_t^\theta) = \mathsf{E}_x^\theta \mathrm{D}f(X_t^\theta)\left(\frac{\partial_\theta X_t}{\mathrm{D}X_t}\right)& \\ = \mathsf{E}_x^\theta f(X_t^\theta)\Xi_t^1 = \mathsf{E}_x^\theta f(X_t^\theta) g_t^\theta(x, X_t^\theta);&\end{aligned} \tag{36}$$

see (7) for the definition of $g_t^\theta(x,y)$. Because the test function $f$ is arbitrary, the integral identity (36) proves (6).

## 5. Proof of Theorem 2

Let us give first an auxiliary result, which gives moment bounds for $\Xi_t, \Xi_t^1$ from Theorem 1 and some consequent properties of $p_t^\theta$ and $g_t^\theta = \partial_\theta \log p_t^\theta$.

**Lemma 2.** *1. For every $p \geq 1$ there exist constant $C$ which depends on $t$ and $p$ only, such that*

$$\mathsf{E}_x^\theta\Big|\Xi_t\Big|^p \leq C, \tag{37}$$

$$\mathsf{E}_x^\theta\Big|\Xi_t^1\Big|^p \leq C(1+|x|)^p. \tag{38}$$

*2. For every $p \geq 1$ there exists constant $C$ which depends on $t$ and $p$ only, such that*

$$p_t^\theta(x,y) \leq C(1+|x-y|)^{-p}. \tag{39}$$

*3. For every $p \geq 1$ there exists constant $C$ which depends on $t$ and $p$ only, such that*

$$\mathsf{E}_x^\theta\Big|g_t^\theta(x, X_t)\Big|^p \leq C(1+|x|)^p. \tag{40}$$

*Proof.* 1. We prove (38), the proof of (37) is similar and simpler. Let us analyse the moment properties of the terms involved into (35). By (26), (29), and the standard properties of stochastic integrals, for any $p \geq 1$ there exists a constant $C^1_{t,p}$ such that

$$\mathsf{E}^\theta_x |\mathrm{D}X_t|^p + \mathsf{E}^\theta_x |\mathrm{D}^2 X_t|^p \leq C^1_{t,p}.$$

By Lemma 1, for any $p \geq 1$ there exists a constant $C^2_{t,p}$ such that

$$\mathsf{E}^\theta_x |\mathrm{D}X_t|^{-p} \leq C^1_{t,p}.$$

Finally, the linear growth condition (5) and (30), (31) yield via the Gronwall lemma that for any $p \geq 1$ there exists a constant $C^3_{t,p}$ such that

$$\mathsf{E}^\theta_x |\partial_\theta X_t|^p + \mathsf{E}^\theta_x |\mathrm{D}(\partial_\theta X_t)|^p \leq C^3_{t,p}(1+|x|)^p.$$

Combining all the above, we get the required bound by the Hölder inequality.

2. The argument here is well known, e.g [11], Chapter 3.1, hence we just sketch it. It follows from the representation (4) and the moment bound (37) with $p = 2$ that

$$p^\theta_t(x,y) \leq C \mathsf{P}^{1/2}_\theta (X_t > y).$$

Recall that $Z_t$ has finite moments of every order and $a_\theta$ has bounded derivative in $x$. Then a standard argument based on the Gronwall lemma provides that

$$\mathsf{E}^\theta_x |X_t - x|^p \leq C$$

with the constant $C$ which depends on $t$ and $p$ only. Then for $y > x$

$$\mathsf{P}_\theta(X_t > y) \leq \min\{1, C|x-y|^{-p}\},$$

which gives a required bound. For $y < x$ one should use instead of (4) the representation

$$p^\theta_t(x,y) = -\mathsf{E}^\theta_x \left[\Xi_t \mathbf{1}_{X_t \leq y}\right],$$

which is equivalent to (4) because $\Xi_t$ is a stochastic integral and therefore has zero expectation.

3. Because

$$g^\theta_t(x, X_t) = \mathsf{E}^\theta_x \left[ \Xi^1_t \middle| X_t \right],$$

the required statement follows from (38) and the Hölder inequality. □

Let us proceed with the proof of the Theorem. We fix $T$ in the constructions from Section 3 greater than $t_n$.

Formula (10) and statement I of Theorem 1 immediately provide the continuity property (a) from the definition of a regular statistical experiment. To prove the $L_2$-differentiability property (b), we put for $\varepsilon > 0$

$$\psi_\varepsilon(z) = \begin{cases} 0, & z < \varepsilon/2, \\ \frac{z-\varepsilon/2}{\varepsilon^{3/2}}, & z \in [\varepsilon/2, \varepsilon], \\ \sqrt{z} - \frac{7\sqrt{\varepsilon}}{8}, & z \geq \varepsilon. \end{cases}$$

By the construction, $\psi_\varepsilon \in C^1$ and $\psi_\varepsilon(z) = 0$ for $z \leq \varepsilon/2$. Then, by statement II of Theorem 1 and statement 2 of Lemma 2 the mapping $\theta \mapsto \eta^\theta_\varepsilon := \psi_\varepsilon(p^\theta) \in L_2(\mathbb{R}^n, \lambda)$ is continuously differentiable with the derivative equal

$$\zeta^\theta_\varepsilon := \frac{d}{d\theta}\eta^\theta_\varepsilon = \psi'_\varepsilon(p^\theta)\partial_\theta p^\theta;$$

see (11) for the formula for $\partial_\theta p^\theta$. By the construction,

$$\psi_\varepsilon(z) \to \psi_0(z) := \sqrt{z}, \quad \psi_\varepsilon'(z) \to \psi_0'(z) = \frac{1}{2\sqrt{z}}, \quad \varepsilon \to 0.$$

Hence to prove the $L_2$-differentiability property (b) it is enough to show that

$$\eta_\varepsilon^\theta \to \eta_0^\theta := \psi_0(p^\theta), \quad \zeta_\varepsilon^\theta \to \zeta_0^\theta := \psi_0'(p^\theta)\partial_\theta p^\theta \quad \text{in} \quad L_2(\mathbb{R}^n, \lambda) \tag{41}$$

uniformly by $\theta$. We show the second convergence in (41), the proof of the first one is similar and simpler. Then by the explicit form of $\psi_\varepsilon'$ we get using the Cauchy inequality

$$\int_{\mathbb{R}^n} (\zeta_\varepsilon^\theta - \zeta_0^\theta)^2 \mathrm{d}\lambda \le \frac{1}{4}\int_{p^\theta \le \varepsilon} (g^\theta)^2 p^\theta \mathrm{d}\lambda \le \frac{1}{4}\left(\int_{\mathbb{R}^n} (g^\theta)^4 p^\theta \mathrm{d}\lambda\right)^{1/2} \left(\int_{p^\theta \le \varepsilon} p^\theta \mathrm{d}\lambda\right)^{1/2}.$$

By statement 3 of Lemma 2 with $p = 4$ we have

$$\int_{\mathbb{R}^n} (g^\theta)^4 p^\theta \mathrm{d}\lambda = \mathsf{E}_x^\theta \left(\sum_{k=1}^n g^\theta_{t_k - t_{k-1}}(X_{t_{k-1}}, X_{t_k})\right)^4$$
$$\le n^3 \sum_{k=1}^n \mathsf{E}_x^\theta \left(g^\theta_{t_k - t_{k-1}}(X_{t_{k-1}}, X_{t_k})\right)^4 \le n^3 C \sum_{k=1}^n \mathsf{E}_x^\theta \Big(1 + |X_{t_{k-1}}|\Big)^4 \le \tilde{C},$$

where constant $\tilde{C}$ does not depend on $\theta$. On the other hand,

$$\int_{p^\theta \le \varepsilon} p^\theta \mathrm{d}\lambda \le \sqrt{\varepsilon}\int_{\mathbb{R}^n} \sqrt{p^\theta}\mathrm{d}\lambda,$$

and by statement 2 of Lemma 2 with $p > 2$

$$\int_{\mathbb{R}^n} \sqrt{p^\theta}\mathrm{d}\lambda \le \hat{C}$$

with constant $\hat{C}$ which does not depend on $\theta$. Summarizing all the above, we get the second convergence in (41), uniform by $\theta$. This completes the proof of the regularity. The formula for the Fisher information follows from the identity

$$\int_{\mathbb{R}^n} \left(\partial_\theta \sqrt{p^\theta}\right)^2 \mathrm{d}\lambda = \frac{1}{4}\int_{\mathbb{R}^n} (g^\theta)^2 p^\theta \mathrm{d}\lambda = \frac{1}{4}\mathsf{E}_x^\theta \left(\sum_{k=1}^n g^\theta_{t_k - t_{k-1}}(X_{t_{k-1}}, X_{t_k})\right)^2,$$

Markov property, and the observation that

$$\mathsf{E}_x^\theta g_t^\theta(x, X_t) = 0$$

for every $x \in \mathbb{R}, \theta \in \Theta, t > 0$, which follows from (36) with $f \equiv 1$.

## References

[1] K. Bichteler, J.-B. Gravereaux, J. Jacod. *Malliavin calculus for processes with jumps.* Gordon and Breach science publishers, N.Y., London, Paris, Tokyo, 1987.

[2] J.-M. Bismut. Calcul des variations stochastique et processus de sauts. *Z. Warw. theor. verw. Geb.*, 56(4):469-505, 1981.

[3] P. Carr, H. Geman, D.B. Madan, M. Yor. Stochastic volatility for Lévy processes. *Math. Finance*, 13:345382, 2003.

[4] L. Chaumont, G. Uribe Bravo. Markovian bridges: Weak continuity and pathwise constructions. *Ann. Probab.*, 39(2):609-647, 2011.

[5] J.M. Corcuera, A. Kohatsu-Higa. Statistical inference and Malliavin calculus. *Seminar on Stochastic Analysis, Random Fields and Applications VI, Progress in Probability, Springer Basel*, 63:59-82, 2011.

[6] E. Fournié, J.M. Lasry, J. Lebuchoux, P.-L. Lions and N. Touzi. Some applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3:391-412, 1999.

[7] I.I. Gikhman, and A.V. Skorokhod. *Stochastic differential equations*. Naukova Dumka, Kiev, 1968. In Russian.

[8] I. A. Ibragimov, R. Z. Hasminskii. *Statistical estimation: asymptotic theory*. New York, Springer-Verlag, 1981.

[9] A.M. Kulik. Absolute continuity and convergence in variation for distributions of functionals of poisson point measure. *Th. Pr.*, 24:1-38, 2011.

[10] E.A. Novikov. Infinitely divisible distributions in turbulence. *Phys. Rev. E*, 50:R3303R3305, 1994.

[11] D. Nualart. Analysis jn Wiener space and anticipating stochastic calculus. *Springer Berlin / Heidelberg, Lecture Notes in Mathematics*, 1690:123-220, 1998.

[12] J. Rosiński. Tempering stable processes. *St. Pr. and their App.*, 117(6):677-707, 2007.

[13] B. Simon, M. Read. *Methods of Modern Mathematical Physics*. Functional Analysis, Academic Press, San Diego, 1972.

[14] T. Simon. Support theorem for jump processes. *Stoch. Proc. and Appl.*, 89:1-30, 2000.

[15] A.V. Skorokhod. On differentiability of the measures which correspond to stochastic processes. I. Processes with independent increments. *Prob. Theory and Appl.*, 2(4):417-443, 1957. In Russian.